

patient data. The method comprises receiving a medical report for a patient including patient identification data. A patient record is searched for an APID corresponding to the patient. The search returns the APID in response to locating it and returns a null value in response to not locating an APID. If the search returns a null value, an APID corresponding to the patient is created. The creating includes receiving a media access control (MAC) address and applying a first linear transformation matrix to the MAC address, resulting in a transformed MAC address. The creating also includes receiving a date/time and applying a second linear transformation matrix to the date/time, resulting in a transformed date/time. Further, the creating includes receiving an anonymity supplement and applying a third linear transformation matrix to the anonymity supplement, resulting in a transformed anonymity supplement. The transformed MAC address, transformed date/time and transformed anonymity supplement are concatenated resulting in the APID. Finally, the creating includes encrypting the APID and storing the encrypted APID in the patient record. The method for creating anonymity in collecting patient data further comprises adding the APID to the medical report and the patient identification data is removed from the medical report. The medical report is transmitted to a data repository in response to removing the patient identification data.

[0006] Another aspect of the invention is a system for creating anonymity in collecting patient data. The system comprises a network and a host system in communication with the network. The host system includes software to implement a method comprising receiving a medical report for a patient including patient identification data. A patient record is searched for an APID corresponding to the patient. The search returns the APID in response to locating it and returns a null value in response to not locating an APID. If the search returns a null value, an APID corresponding to the patient is created. The APID is added to the medical report and the patient identification data is removed from the medical report. The medical report is transmitted to a data repository in response to removing the patient identification data.

[0007] A further aspect of the invention is a computer program product for creating anonymity in collecting patient data. The computer program product comprises a storage medium readable by a processing circuit and storing instructions for

collecting patient medical records generated by a medical site, such as a hospital, for use in public data mining. An embodiment of the present invention can facilitate the mining of medical outcomes and diagnosis and enhance patient care. The collection of this information maintains grouping of patient studies through the creation and use of an anonymous patient identifier (APID). For example, if John Doe had ten studies during one or more visits, then all of the studies for John Doe will be grouped under a single APID. The structure of the APID is designed to prevent duplication with other APIDs generated at the same or other sites using the same process. This is accomplished by creating a unique surrogate identifier, the APID, that is assigned to each patient. When a patient is created (manually or received by another system) in the departmental system, a unique APID is created for that patient. The APID is automatically encrypted and stored with the patient data in the patient record section of the database. The APID is encrypted by a separate application that has no access other than the APID encryption. Each time a patient's study report is copied for collection, the patient identification items (e.g., patient name and patient ID) are removed and replaced by the APID.

[0015] FIG. 1 is a flowchart of an exemplary process for creating anonymity in collecting patient data. At step 102, a report is selected to be sent to the data repository for use in public data mining. Report selection may be triggered when the report is completed, or alternatively, the database of reports may be periodically searched and reports in the database selected for the data repository based on pre-selected search criteria. Search criteria can include things like selecting all reports that have not been previously sent to the data repository or selecting only reports that relate to particular medical problems or treatments. At step 104, a check is made to determine if the patient record associated with the patient specified in the report already includes an encrypted APID that can be linked to the selected report. If the patient record includes an encrypted APID, processing continues at step 112. Otherwise, an encrypted APID must be created to correspond to the patient in the report. At step 106, an APID is generated for the patient. FIG. 2, discussed below, describes an exemplary embodiment of a process for creating an APID. At step 108, the APID is encrypted using any encryption software known in the art (e.g., PGP Corporation's PGP and RSA's BSAFE). At step 110, the encrypted APID is stored on the patient record.

[0016] Next, at step 112, the encrypted APID associated with the patient in the report selected for collection for the data repository is unencrypted using decryption software that corresponds to the encryption software utilized at step 108. In an exemplary embodiment, if the patient record did not include an encrypted APID, the unencrypted APID created in step 106 could be utilized and step 112 could be skipped. At step 114, patient identification data is removed from the report. Patient identification data includes any information or combination of information that could be used to identify a specific individual and can include name, social security number, insurance numbers and address. At step 116, the unencrypted APID is added to the report and at step 118, the report is sent to the data repository for use in public data mining.

[0017] FIG. 2 is a flowchart of an exemplary embodiment for creating an APID 218 for use in an embodiment of the present invention. It is important that a unique APID 218 is created for each patient across all hospitals and data sources for the data repository. In an exemplary embodiment of the present invention the APID 218 includes three components: a media access control (MAC) address; a date/time and an anonymity supplement. As depicted in FIG. 2, the MAC address 202 is input to a first linear transformation matrix 204. The MAC address 202 is a unique number that is burned into an Ethernet or token ring adapter that distinguishes it from all other network cards. The use of the MAC address 202 can ensure uniqueness among the mutually exclusive systems, or hospitals, that may be utilized to collect data for the data repository. The linear transformation matrix can be any non-singular linear matrix. In an exemplary embodiment the non-singular matrix is a three by three matrix, in another it is a one by three matrix. Using a three by three matrix as an example, the defined linear transformation is:

$$L:R^3 \rightarrow R^3 \text{ by } L(X) = AX$$

The output from the linear transformation matrix applied to the MAC address 202, the transformed MAC address, is denoted as M' in FIG. 2.

[0018] The second component to the APID 218 is the date/time 206 and it represents the date and time that the patient record was created in the system. In an exemplary embodiment, the date/time 206 component includes three subcomponents: date in "mmddyy" format, time in "hhmmss" format, and a constant digit. The date/time 206

component is utilized in order to provide uniqueness within the hospital where the patient was treated. The date/time 206 component also goes through a second linear transformation matrix 208 to prevent the ability to guess all of the patients for a given date and time. This second linear transformation matrix 208 may be the same or different than the first linear transformation matrix 204 applied to the MAC address 202. In an exemplary embodiment, the second linear transformation matrix 208 can be any non-singular linear matrix such as a three by three matrix or a four by four matrix. Using a three by three matrix as an example, the defined linear transformation is:

$$L:R^3 \rightarrow R^3 \text{ by } L(X) = AX$$

The output from the linear transformation 208, the transformed date/time, is denoted as D' in FIG. 2. Both M' and D' are input to a concatenate function 210 to be concatenated together and the resulting output is denoted as M'D'.

[0019]

The third component to the APID 218 is the anonymity supplement 212 component. The anonymity supplement 212 component includes three subcomponents: a random number between 1 and 1000, a rotating number between 20 and 40, and a constant. The rotating number subcomponent increments by 1 each time an anonymity supplement is created and once it gets to 40 it returns back to 20. The anonymity supplement 212 component is utilized in order to prevent the APID from being predicted by individuals trying to identify the individual patient being discussed in the report. It also is utilized to get around the case where two patients are admitted to the same hospital at exactly the same time. The anonymity supplement 212 component also goes through a third linear transformation matrix 214. This third linear transformation matrix 214 may be the same or different than the linear transformation matrices 204 208 applied to the MAC address 202 and date/time 206 components of the APID 218. In an exemplary embodiment, the third linear transformation matrix 214 can be any non-singular linear matrix such as a three by three matrix or a four by four matrix. Using a three by three matrix as an example, the defined linear transformation is:

$$L:R^3 \rightarrow R^3 \text{ by } L(X) = AX$$

The output from the third linear transformation matrix 214, the transformed anonymity supplement, is denoted as S' in FIG. 2. Both M'D' and S' are input to a

concatenate function 216 to be concatenated together and the resulting output is denoted M'D'S' which is the APID 218.

[0020] The APID 218 is then sent through an encryption program 220 to create an encrypted APID 222 that is then stored on the patient record 224. When a request is made to attach an APID 218 to a report, a decryption program 226 that corresponds to the encryption program 220 is utilized to transform the encrypted APID 222 into an unencrypted APID 218. The unencrypted APID 218 is then stored on the report or linked to the report that is sent to the data repository for public data mining. The linear transformation matrices described above can differ between hospitals and are typically stable for ease in transformation. In an exemplary embodiment, the first linear transformation matrix 204 utilized on the MAC address 202 is the same for all implementations and the other linear transformation matrices 208 214 differ between hospitals and are modified by product upgrades.

[0021] FIG. 3 depicts the transformation of a data stream into an exemplary encoded APID 218. The data contained in box 302 is a sample MAC address 202, a sample first linear transformation matrix 204, "LT", and the transformed MAC address, M'. Box 304 includes a sample date/time 206, a sample second linear transformation matrix 208, "LT", and the transformed date/time, D'. Box 306 includes a sample anonymity supplement 212, a sample third linear transformation matrix 214, "LT", and the transformed anonymity supplement, S'. Box 308 contains the resulting APID 218 which is created by concatenating the three values: M', D' and S' together. Box 310 contains an example of an encrypted APID 222. The data values and formats depicted in FIG. 3 are meant to be examples of one way to implement the creation of an APID 218 utilizing the present invention, each implementation will include different linear transformation matrices and may include different data formats for the date/time 206 and anonymity supplement 212 components. Further, the MAC address is utilized to uniquely identify a computer system and any identifier that also uniquely identifies a computer system may be used in place of the MAC address. In addition, any encryption software known in the art can be utilized with an embodiment of the present invention.

[0022] An embodiment of the present invention allows for the creation of a unique and

anonymous patient identifier that is attached to a medical report for use in grouping reports relating to a single patient without revealing the identity of the patient. This can allow for a secure mechanism for gathering anonymous patient medical data for use in public data mining. By including the MAC address, an existing and unique alphanumeric value assigned to each system, in the APID 218, different mutually exclusive systems can be utilized to originate anonymous patient medical data without resulting in duplicate APIDs 218 between systems. The use of the date/time 206 and the anonymity supplement 212 helps to ensure that the APIDs 218 are not duplicated within a single system. The ability to create a unique APID 218 can allow for more meaningful data mining because reports relating to a single patient can be reviewed and analyzed as a group. The ability to prevent a public data mine user from tracing a report back to an individual is important to protecting patient privacy. This is accomplished by sanitizing the medical reports by replacing the patient data with an APID 218 before transmission to a data repository and by creating an APID 218 that can not be traced back to an individual patient. Providing public data mining access to data reports, grouped by individual patient, can lead to better information for use by pharmaceutical companies and hospitals in improving medical procedures and products.

[0023]

As described above, the embodiments of the invention may be embodied in the form of computer-implemented processes and apparatuses for practicing those processes. Embodiments of the invention may also be embodied in the form of computer program code containing instructions embodied in tangible media, such as floppy diskettes, CD-ROMs, hard drives, or any other computer-readable storage medium, wherein, when the computer program code is loaded into and executed by a computer, the computer becomes an apparatus for practicing the invention. An embodiment of the present invention can also be embodied in the form of computer program code, for example, whether stored in a storage medium, loaded into and/or executed by a computer, or transmitted over some transmission medium, such as over electrical wiring or cabling, through fiber optics, or via electromagnetic radiation, wherein, when the computer program code is loaded into and executed by a computer, the computer becomes an apparatus for practicing the invention. When implemented on a general-purpose microprocessor, the computer program code

